The Application of JASP Data Visualization in Blended Foreign Language Teaching

Ran Li *

School of English Language and Literature, Xi'an Fanyi University, Xi'an 710105, China

542501525@qq.com

*corresponding author

Keywords: data visualization, blended foreign language teaching, data mining

Abstract: As the philosophy of foreign language teaching constantly evolves in the direction of precision, informatization and intelligence, the corresponding teaching modes also transforms with the aid of information and internet technology, advocating the precise teaching design, teaching evaluation and the data-driven decision-making in teaching. Based on the statistical and visualization module of JASP, this article explores the application of data visualization software like JASP in blended foreign language teaching, shedding light on the visualization and statistical analysis of qualitative data accumulated in four rounds of blended teaching, highlighting the significance and value of data analysis and data visualization techniques for the blended teaching of foreign languages in the age of intelligence.

1. Introduction

Recently, with the popularization of Artificial Intelligence and Big Data technology in a growing number of fields, the field of foreign language education is turning to data mining and data visualization to make sense of a sea of education data. Data visualization can be used to address a wide range of practical problems in blended foreign language teaching, such as the evaluation of students' learning situation, creating student's academic performance portfolios and optimizing blended teaching design.

Data visualization is critical to understanding the nature of the formative data acquired from regular observation of teaching and learning activities. Aided by software, foreign language teacher can become data analysts using visualization techniques to examine, scrutinize, and validate their analysis of student's learning situation and status, the effect of their own teaching, to pave the way for personalized precision teaching. Using data mining models designed specifically for foreign language teaching foreign language teachers derive data sets from the statistical summarizations of the teaching and learning process. These data sets can be used to generate data graphics which can help us think visually and provide a powerful way to reason about the blended teaching and learning process.

The most desirable way to visualize the multidimensional data sets acquired from the blended foreign language teaching is to describe the items by more than three features which are input as variables, so as to make some visual insight into the practical problems in blended foreign language teaching.

All in all, the principal goal of data visualization analysis in blended foreign language teaching is to locate, identify, discriminate, categorize, cluster the data sets and determine the pattern of its distribution, ordering, correlation, which can greatly enhance teachers' cognitive efficiency and guide them through the inference of valid information from the results of visualized analysis of data sets.

2. The Visualization Analysis of the Empirical Data from the Blended Teaching of Advanced English

In the present study, JASP is adopted as the comprehensive learning analysis tool. JASP is developed by professor Eric-Jan Wagenmakers and his team in the University of Amsterdam. JASP provides graphic interface, allowing its users to apply multiple statistical analysis such as T-test, variance analysis, covariable analysis, credibility, regression, factor analysis, cluster analysis, integration analysis, etc. Users may select different analytic functions to make personalized adjustments to parsing algorithm.

Utilizing the data-fetching and statistical analysis functions of the on-line course platform, the author monitors the online learning status and online learning process of three classes for a semester, collects relevant qualitative data and imports the data into JASP for data visualization, which manifests the empirical validity of data visualization in the measurement of the effects of blended foreign language teaching.

2.1. Participants

The participants are junior English majors at a private college. They specialized in translation and have participated in the blended teaching program for 2 semesters, accordingly, they have been acquainted with the on-line curriculum provision of *Advanced English* and can follow the on-line learning and teaching program actively, accomplishing the online learning tasks timely.

2.2. Data Collection

Advanced English has built up mature Provincial High-quality on-line course system on Xueyin online course system, in the blended teaching practice of school-based curriculum, massive online and offline data of the corresponding teaching and learning practice. Based on the predetermined on-line observation points, teachers record and evaluate students' leaning situation, learning aptitude and learning needs via the surveying and mapping technology of precision teaching. At first, teachers establish the evaluation index system, then conduct the multi-dimensional observation and multivariate data collection of student's learning situation, decoding students' learning needs. Afterwards, teachers utilize the statistical and analytic tool embedded in the online platform, such as the statistical module and data analysis module and data visualization module, to visualize the data pertaining to the online learning situation, generating learning report, graphs on the distribution of knowledge points and achievement trend etc.

Moreover, teachers devise the standard deviation charts to assess the offline learning situation,which will be integrated with the results of the online learning situation data analysis to

further delineate learners' cognitive and psychological traits, cognitive style, learning style, learning needs and learning status, to detect the common and individual problems. Targeted at the above-mentioned problems, teachers make plans for precision teaching and implement personalized and differentiated teaching.

In the present study, most of the online qualitative data is obtained from following online observation points: online learning activity participation, such as the frequency of online interaction; online learning process, such as the visit traffic and the access time of course task points and teaching resources; online learning results, such as the completion of online homework, online tasks and group cooperative learning tasks (PBL) and online test score.

Similarly, the offline observation points focus on students' academic performance, such as the record of classroom learning process, the audio and video records of the classroom management, reflecting the patterns of teachers classroom discourse, and students' questioning and answering, as well as the coordination of classroom activities and interactions. The rest of the offline observation concerns students' learning results, including scores of pre-assessment and post-assessment, records of classroom interaction and group learning activities, such as the learning accomplishments cards, learning notes and reflections, written homework and the completion of group cooperative learning tasks.

2.3. The Mining of Formative Data

The data recorded on the Xueyin online course platform can be divided into two categories: one is the structured data, stored in the form of two-dimensional table; the other is the unstructured data, such as the text message stored in the discussion boards and evaluation area, including the records of online discussion, the content and number of posts (texts, photos and accessories). The acquisition of data recorded on the learning platform is also the process of data mining of the teaching and learning process, which involves statistical means, text mining methods and natural language processing methods. The author starts the longitudinal action research on the effects of blended teaching of Advanced English framed by the BOPPPS model since September, 2020, in which the author documented and collected the online learning data of 120 students from 3 classes. The process of data collection spans a semester to December, 2020 and the online data will be exported monthly as the basis for formative evaluation, with four Excel files for each class.

The online formative data collection chart students' online learning trajectory via the data fetching of online platform, including the frequency of signing-in, the learning of chapter knowledge points, the participation of online discussion, online survey, online classroom activities and the completion of PBL tasks. Meanwhile, the learning status of learners will be assessed via the statistical records of online data, including the completion of the task points of basic and extended course resource, online homework score, online test score and the integral of online course interaction and teachers' ratings of PBL projects. Eventually, all the online data can be extracted as individual academic performance report for learning feedback and supervision.   The extracted data encompass fields such as personal information, course video learning score, the learning progress of course resources, online discussion score, online homework score, online test score, and online interaction score etc.

The offline formative data mainly comes from the standard deviation charts which document students' offline learning behavior in accordance with the preset formative evaluation rubrics, the indexes include group cooperative learning score obtained from intra-group and inter-group evaluation, teacher's rating of group cooperative learning results and individual inquiry learning

results. Furthermore, summative evaluation methods like final examination will also be adopted to assess students' lower-order thinking abilities, as well as the higher-order cognitive and thinking abilities.

2.4. Standardizing the Variables and Coding Student Data

In the data processing stage, the online structured data will be integrated with the offline formative data. According to the preset formative evaluation standards, the author extracts the online structured data and assign values to fields such as online homework, online test, audio and video learning, chapter learning, online discussion, course interaction. All the values will be imported into the offline formative evaluation chart via the VLOOKUP function of Excel to be integrated with the filed values of the corresponding evaluation indexes. To enhance the efficiency of data processing, the monthly evaluative data of each class will be integrated into one Excel file. The null value of fields caused by the unsubmitted homework or unattended tests will be replaced by the integer zero.

In order to treat different variables with the same weight, a scheme for normalizing the variables and coding student data has been devised. As an example, the original data set containing personal information variables stored as Chinese characters, such as major, college, student ID, class, gender etc, can not be identified and processed normally in JASP. Therefore, such data has to be standardized and substituted with English expression or numerals. Accordingly, students' names will be abbreviated and coded as pinyin initials, the rest fields of evaluative indexes will be substituted with English expression to ensure the consistent treatment of corresponding variables in JASP. All the online and offline evaluative indexes will be input as the attributes or variables in data analysis and each Excel file has 21 attributes or variables.

The data visualization analysis in this study integrates the valid data of 116 students from 3 classes, to adapt to the requirements of JASP, the above-mentioned Excel file in .xls format will be converted into .csv format, finally outputting the following evaluation chart:

	Name	Online Ho	Online Ho	Online Ho	Online Ho	Online Te	Audio Vid	Chapter L	Online Di	PBL		Online Cl	Absence	Leave	Score	Offline Ho	Offline H	Off
2	DHY	100	95	0	100	100	0	6.6	0	90		11	1	1	3	95	0	
3	SSW	100	100	99	100	95	9.6	19.4	0	96		16			5	95	94	
4	WXY	100	75	94	100	100	0	6.6	0	90		14			5	96	93	
5	YYN	100	100	97	100	95	0	13	2	90		16			5	93	95	
6	LT	100	100	0	100	100	0.6	9	0	96		19			5	96	94	
7	CSR	100	100	0	100	95	5.4	18.6	0	96		16			5	96	0	
8	ZYX	90	95	77	90	100	0	8.4	0	95		14			5	94	95	
9	GLY	100	100	0	100	100	6	21	0	95		14			5	95	0	
10	ZL	50	85	77	100	95	0	8	0	96		14			5	92	93	
11	WXQ	100	100	78	100	100	5.4	26.4	0	95		11			5	94	95	
12	ZNN	100	95	72	100	0	9.6	20	0	95		18			5	94	96	
13	JXY	80	100	77	90	95	9.6	57.6	2	90		16			5	93	97	
14	FXR	100	100	51	100	100	12	22	0	95		11			5	96	95	
15	HSY	100	100	94	100	95	0	16	0	90		14			5	93	92	
16	ZYX	100	100	95	100	100	5.4	27.6	0	90		14			5	94	94	
17	WJL	100	100	94	100	100	0	12.4	0	90		14			5	94	94	
18	GYX	100	100	0	100	100	0	3	0	96		14			5	95	0	
19	ZX	100	100	98	100	100	7.8	29	0	95		16			5	96	94	
20	ZJM	80	100	65	100	100	3	16	0	95		14			5	90	96	
21	RYW	90	100	0	75	5	3	20	0	95		16			5	94	97	
22	CXY	100	100	88	100	85	9.6	34	0	95		16			5	94	95	
23	XYY	100	100	0	100	90	7.2	15.6	0	95		16			5	94	95	

Figure 1: the standardized evaluation chart

3. Results and Discussion

Using the descriptive statistical analysis module of JASP, the author conducted the visualization analysis of the 21 attributes in the formative evaluation chart. The comparative visualization

analysis of the monthly formative data from September to December clearly reveals students' learning effects and the effects of blended foreign language teaching.
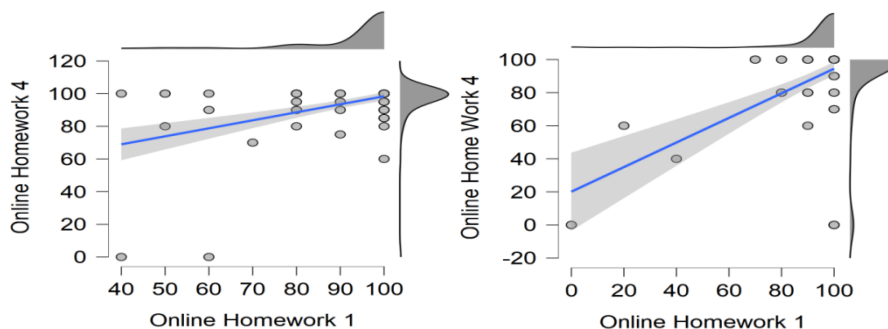
3.1. Online Homework

Figure 2: The visualization of the scores of online homework

The degree of completion and the score of online homework is the essential index in the assessment of students' online learning status. The present analysis visualized the scores of the first online homework in September and the fourth online homework in December and made a frequency analysis. The frequency table generated by JASP suggests, among the 116 valid sample data, the percentage of 100 points has increased from the original 77.5% to the present 81.8%. Meanwhile, in the descriptive statistical analysis of JASP, the scores of the first and fourth online homework in September and the scores of the first and fourth online homework in December are used as variables in two rounds of regression analysis. The X axis is the value field of the score of the first online homework; the Y axis is the value field of the score of the fourth online homework. The data distribution in the left scatter plot is more discrete, indicating the correlation between the two variables is not very prominent, and namely, the score of the first online homework in September as the independent variable has no apparent correlation with the dependent variable, that is, the score of the fourth online homework. By contrast, the regression line on the scatter plot of December to the right slopes upper right more apparently, suggesting the clustering of data distribution, which further leads to the conclusion that the scores of the two online homework have the positive correlation.

The above visualization analysis suggests the stability of score correlates positively with the completion of online homework, that is, the consistent online learning behavior.
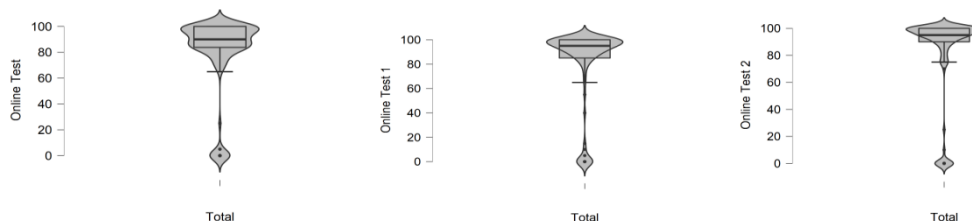
3.2. Online Test

Figure 3: The visualization of the online test score

Violin Plot is one of visualization models in JASP which is used to present the distribution status and probability density of multiple data sets. The present study conducts data modelling via the descriptive statistical analysis of the scores of the first online test in September and the scores of two online tests in December, generating three violin plots in Figure 3, which clearly visualize the variation in students' online test score from September to December.

The distribution of data shows a gradual dense clustering from 60 to 80, then towards 100, suggesting the increase in the number of students whose score ranges from 80 to 100, also the polarization of students' online test score.
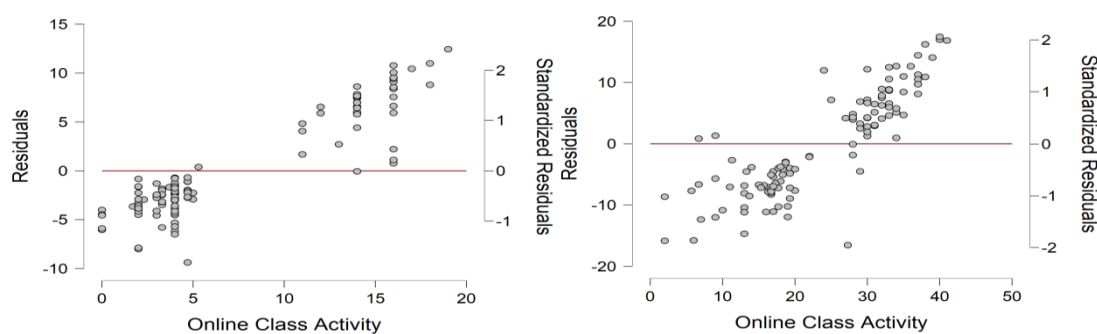
3.3. Online Class Activity

Figure 4: The visualization of online activity

The online activities adopted in the blended teaching of Advanced English include online survey, in-class activities, random selection of students, voting, preemptive answers, topic discussion, which will be integrated with in-class teaching activities to mobilize the students. In JASP, the present study conducted unitary linear regression analysis of the relevant data of the online class activities in September and December. The online class activity score is imported as dependant variable, the scores of video and audio learning, chapter learning and online discussion are imported as covariables to generate the scatter plot in figure 4 with the axis X as the value field of online classroom activity, axis Y as the residuals. As is indicated by the data distribution characteristics in figure 4, from September to December, students' participation in online class activity has increased prominently, as well as the frequency of other online learning activities.

3.4. Group Cooperative Learning Projects (PBL)

In the blended teaching of Advanced English, teachers assign cooperative learning tasks in learning guidance. The task-based group cooperative learning results in the form of papers, report or video and audio summary will be uploaded onto the online platform by students and teachers will evaluate and rate according to the preset standards.
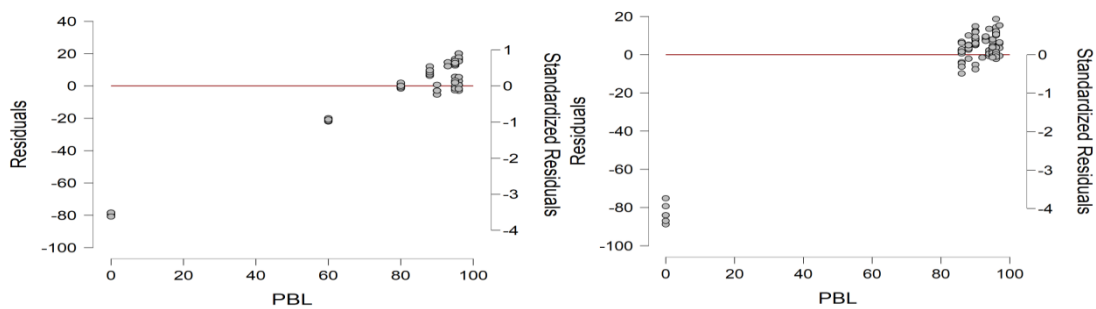
Figure 5: the visualization of the scores of Group Cooperative Learning Tasks

In the unitary linear regression analysis, the scores of group learning tasks in September and December are imported into JASP with the scores of PBL as dependent variable and the scores of online class activities as covariable. From left to right, the results of the visualization analysis of the scores of September and that of December is shown in the above scatter plot, with the axis X as the value field and axis Y as the residual. It is quite apparent in the graph that the tendency of clustering is more prominent in the scatter plot of December, which suggests the frequency of students' participation in group learning projects has greatly enhanced.

3.5. Offline Homework

The quality of offline homework is also a very important index of the effects of blended teaching. As the indispensable formative data, the score of offline homework may attest to students' learning effects in the course of participatory leaning in blended teaching.
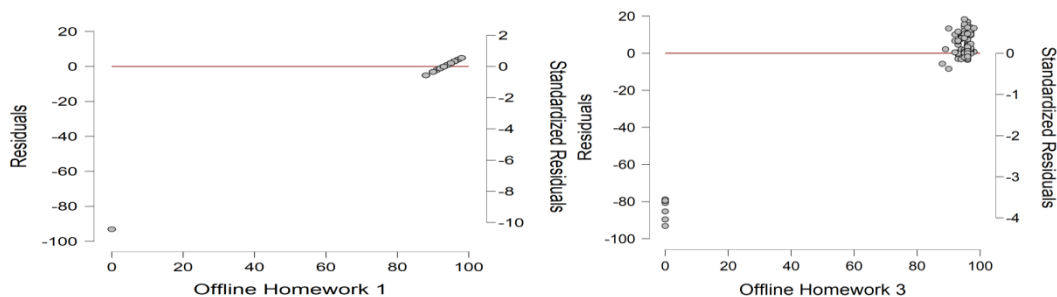

Figure 6: the visualization of the scores of offline homework scores

In the blended teaching design of Advanced English, student do offline homework three times a month, which could be the analysis and paraphrase of long and difficult sentences, translation practice and dictation etc. In JASP, the unitary linear regression analysis takes the score of the first offline homework in September and the score of the third offline homework in December as the dependent variables, with the intra-group evaluation of group learning tasks as the covariable. From left to right, the scatter plots show the data distribution pattern of September and December respectively, with the axis X as the value field and axis Y as the residual. Obviously, the variation in the distribution pattern suggest the frequency of completion and score has increased significantly, especially under teacher's supervision.

3.6. Scores of Monthly Assessment

In the blended teaching of Advanced English, a monthly vocabulary and comprehension test will be released as pre-assessment. At the end of each month, points will be assigned to all the formative evaluation indexes stipulated as the previously-mentioned observation points as post-assessment and summary. The final score at the end of the term will be the average value of the scores of the four months in the term. The variation in the scores of the periodical tests from September to December clearly manifests the effects of blended teaching and learning.

Similarly, the score of monthly assessment will be imported as the dependent variable and the final results of students' average academic performance will be imported as covariable. After the unitary linear regression analysis in JASP, four scatter plots in graph 6 have been generated. The axis X is the value field of monthly assessment and axis Y is the residual.
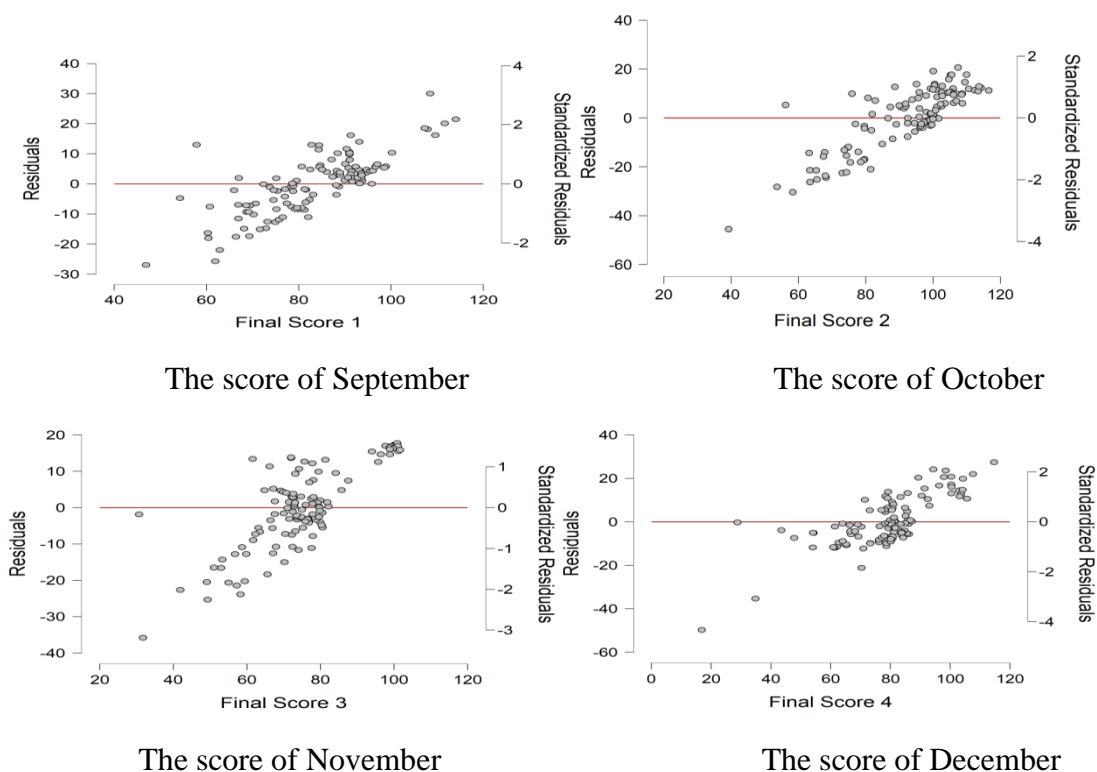

The score of September   The score of October

The score of November   The score of December

Figure 7: The visualization of monthly assessments

The data distribution pattern in the scatter plots indicate: students' scores of the monthly assessment in September vary from 60 to 80. In October, the distribution of data sample shows the tendency of clustering toward the grade level of 80 to 100. In November, students' scores fluctuate in a small range, still, some students' scores steadily rise to the grade level of 100 to 120. In December, most students' scores cluster around the grade level of 80 to 120 and a small portion of students' scores rise moderately. What is noticeable is the scores of some students which originally situate at the grade level of 40 to 60 have increased drastically, clustering around the grade level of 60 to 80.

4. Implications

The paper present a tentative paradigm for the visualization analysis in blended foreign language teaching aid by JASP. The empirical data is drawn from the small-scale qualitative action research on the effects of blended foreign language teaching framed by differentiated formative evaluation design, multi-dimensional observation and analysis of students' learning situation and the precise and personalized curriculum standard and formative evaluation index system. The results of the visualization analysis demonstrates the crucial steps in the multi-modal data mining and analysis in blended foreign language teaching. The visualization analysis of formative data forebodes the future trend of intelligent teaching and learning in the field of foreign language teaching, namely the in-depth integration between foreign language teaching and software technology, big data technology, data mining and analysis, machine learning, which may enhance teachers' ability of data-driven decision-making.

Acknowledgments

References

[1] Boris, Mirkin. *Core Concepts in Data Analysis: Summarization, Correlation and Visualization* [M]. *Springer-Verlag London Limited, 2011*

[2] Han Ye, Gao Xuesong. *The Review on the Recent Research on Online Foreign Language Teaching: Theoretical basis, Core Concepts and Research Methodology* [J], *Foreign Languages and Their Teaching, 2020 (5): 1-11.*

[3] Yu Chunyan, Guo Jinghua. *The Theory and Practice of MOOC and Blended Teaching* [M]. *Beijing: Tsinghua University Press, 2018.*

[4] Chapelle, C. *Technology-Mediated Language Learning* [A]. In J. Schwieter & A. Benati (Eds.). *The Cambridge Handbook of Language Learning* [C]. *Cambridge: CUP, 2019.*

[5] Chen, C. *Analyzing Online Comments: A Language Awareness Approach to Cultivating Digital Literacies* [J]. *Computer Assisted Language Learning, 2020 (4): 435-454.*

[6] Comas-Quinn, A. *Learning to Teach Online or Learning to Become an Online Teacher: An Exploration of Teachers' Experiences in a Blended Learning Course* [J]. *ReCall, 2011 (3): 218 -232.*

[7] Gimeno-Sanz, A. *Moving a Step Further from "Integrative CALL": What's to come?* [J]. *Computer Assisted Language Learning, 2016 (06): 1102-1115.*

[8] Harasim, L. *Learning Theory and Online Technologies (2nd Edition)* [M]. *New York: Routledge, 2012.*